Elucidating Model Inadequacies in a Cloud Parameterization by Use of an Ensemble-Based Calibration Framework

JEAN-CHRISTOPHE GOLAZ

UCAR Visiting Scientist Program, NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

VINCENT E. LARSON

University of Wisconsin—Milwaukee, Milwaukee, Wisconsin

JAMES A. HANSEN

Naval Research Laboratory, Monterey, California

DAVID P. SCHANEN AND BRIAN M. GRIFFIN

University of Wisconsin—Milwaukee, Milwaukee, Wisconsin

(Manuscript received 28 August 2006, in final form 6 February 2007)

ABSTRACT

Every cloud parameterization contains structural model errors. The source of these errors is difficult to pinpoint because cloud parameterizations contain nonlinearities and feedbacks. To elucidate these model inadequacies, this paper uses a general-purpose ensemble parameter estimation technique. In principle, the technique is applicable to any parameterization that contains a number of adjustable coefficients. It optimizes or calibrates parameter values by attempting to match predicted fields to reference datasets. Rather than striving to find the single best set of parameter values, the output is instead an *ensemble* of parameter sets. This ensemble provides a wealth of information. In particular, it can help uncover model deficiencies and structural errors that might not otherwise be easily revealed. The calibration technique is applied to an existing single-column model (SCM) that parameterizes boundary layer clouds. The SCM is a higher-order turbulence closure model. It is closed using a multivariate probability density function (PDF) that represents subgrid-scale variability. Reference datasets are provided by large-eddy simulations (LES) of a variety of cloudy boundary layers. The calibration technique locates some model errors in the SCM. As a result, empirical modifications are suggested. These modifications are evaluated with independent datasets and found to lead to an overall improvement in the SCM's performance.

1. Introduction

Weather and climate models need to simulate a variety of boundary layer clouds, such as cumulus, stratocumulus, and cumulus-under-stratocumulus. Because these clouds are subgrid scale, they must be parameterized. Such parameterization packages should be general enough to simulate any type of cloudy boundary layer that may develop in a forecast or simulation. The parameterizations contain approximate equations governing the sources and transport of heat, moisture, and momentum. Each of these equations contains several terms, such as the turbulent transport of the quantity of interest. Many of the terms must be modeled approximately and hence contain undetermined parameters.

Such parameters are unavoidable. They control, at a minimum, eddy diffusivity and microphysics. Furthermore, these parameters cannot be derived theoretically from first principles. Rather, they must be fitted to data of some sort, either directly or indirectly. That is, the parameters must be estimated or calibrated (Jackson et al. 2003, 2004; Carrió et al. 2006).

Recent advancements in ensemble-based approaches

Corresponding author address: Jean-Christophe Golaz, NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University, Forrestal Campus, P.O. Box 308, Princeton, NJ 08542.
E-mail: Chris.Golaz@noaa.gov

DOI: 10.1175/2007MWR2008.1

to parameter estimation have resulted in a surge of interest in the numerical weather prediction community and the climate community. For example, ensemble Kalman filter (EnKF) methods (Evensen 1994) have been used for simultaneous state and parameter estimation. Hacker and Snyder (2005) used EnKF to assimilate surface layer observations in a boundary layer model and estimate the moisture availability parameter. Aksoy et al. (2006a,b) performed state and parameter estimation for a two-dimensional sea-breeze model and numerical weather predictions using the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5). Zupanski and Zupanski (2006) proposed a method to estimate model errors using an ensemble data assimilation and state augmentation. EnKF methods have also been applied to parameter estimation in climate models of intermediate complexity (Annan et al. 2005a,b).

Despite this increased interest, this process of calibration has a somewhat sordid reputation in the parameterization community. Although every cloud parameterization is calibrated at least informally as a standalone single-column model, the calibration of cloud parameterizations is barely discussed in the literature (see, however, Emanuel and Žiković-Rothman 1999). The reputation of calibration suffers because one often suspects that calibration has been used to mask structural model errors. A structural error is a type of model deficiency in which there is a misspecification of a term's functional form, not merely a misspecification of a parameter value. This paper argues that the evil here is not calibration per se, but rather model structural error; calibration should not be marginalized, but rather exploited to detect model error.

Two common symptoms of structural error are underfitting and overfitting (Geman et al. 1992; Moody 1994; Wilks 1995).

Underfitting occurs when a model's structure is not rich enough to capture true variability in a dataset. In such situations, calibration techniques can distinguish true structural errors from merely poor parameter values, which are not easy to distinguish otherwise. In one common example, a good fit to a data case cannot be achieved despite calibrating all parameter values. Then the remaining errors must be structural in some sense. Another more subtle situation occurs when no single set of parameter values yields a good fit for all cases in the dataset, even though good parameter values can be obtained for each case separately. In this instance, differences in the parameter values in the separate calibrations can provide clues about the source of structural error. In these situations, calibration does not hide errors, but exposes them.

Overfitting occurs when too many parameters are fitted using too few data. Overfitting may hide structural errors because it may introduce compensating errors between terms. This occurs when, in the process of fitting a model to a limited dataset, erroneous parameter values are inadvertently chosen such that one term cancels structural errors in another. It is nontrivial to detect compensating errors when individual terms are not directly observed, as is often the case. If the structural errors persist undetected, then the overfitted model is unlikely to match other, different datasets. In this situation, the model suffers an undesired loss of generality. However, a means to mitigate overfitting is cross validation against independent datasets.

This paper applies an ensemble parameter estimation technique to a single-column model (SCM) for boundary layer clouds and turbulence. Our two main goals are to 1) detect structural model errors in the SCM and 2) improve the SCM's fit over a broad range of cloud regimes.

The structure of this paper is as follows. In section 2, we outline the SCM that we calibrate. In section 3, we describe our ensemble-based parameter estimation framework. In section 4, we discuss the initial parameter estimation experiments and the model deficiencies revealed by them. In section 5, we propose empirical model modifications and test them with reference large-eddy simulation (LES) datasets. In section 6, we cross validate these modifications using independent datasets.

2. Description of SCM

Our SCM simulates boundary layer clouds and is fully described in Golaz et al. (2002a). Briefly, the SCM is a higher-order turbulence closure model that uses a multivariate probability density function (PDF) to close higher-order turbulence and buoyancy terms. The multivariate PDF represents the horizontal subgrid-scale variability of vertical velocity, temperature, and total moisture. A functional form of the PDF is specified, and for each vertical level and time step, moments for that functional form are predicted (such as mean, standard deviation, etc.), thus allowing the PDF to vary with height and time. The underlying functional form of the PDF is a mixture of two trivariate Gaussians. The shape was determined empirically from both aircraft measurements and LES data by Larson et al. (2002) with further modifications by Larson and Golaz (2005). In Golaz et al. (2002b), the SCM was found to perform satisfactorily over a range of boundary layer regimes

comprising clear convection, trade wind shallow cumulus, shallow convection over land, and nighttime stratocumulus. These regimes were all simulated without any case-specific adjustments or trigger functions. Nonetheless, the SCM contains deficiencies whose sources can be difficult to pinpoint.

The deficiencies became more apparent when the SCM was used to simulate a challenging stratocumulus case, the first research flight (RF01) of the second Dynamics and Chemistry of Marine Stratocumulus (DYCOMS-II) field study (Stevens et al. 2003). RF01 was characterized by a strong inversion at cloud top. This inversion was such that stability criteria proposed by Randall (1980) and Deardorff (1980) would have predicted the cloud to dissipate. However, the cloud was observed to persist. As was the case for some LES models (Stevens et al. 2005), our SCM produced a cloud layer that thinned unrealistically over time. To obtain a more reasonable evolution of the cloud layer (Zhu et al. 2005), we modified the mixing length. The present research was partly motivated by the desire to investigate whether the original SCM's difficulty with the DYCOMS-II RF01 case was a structural model error or instead a suboptimal choice of parameter values.

3. Ensemble-based parameter estimation framework

Our boundary layer cloud SCM, like most others, contains nonlinear dynamics and multiple parameters. Given this complexity, it can be difficult to predict the consequences of a particular change in parameter values. This hampers calibration by hand. To expedite the process of parameter estimation, we develop an *automated* calibration framework.

A number of factors guided our choice of a parameter estimation algorithm:

1) Our single-column model is computationally inexpensive compared to three-dimensional models, and the number of parameters we want to estimate is moderate. Therefore, the efficiency of the parameter estimation algorithm is not an urgent concern.
2) We wish to use a uniform prior parameter distribution, so as to enable the algorithm to yield the aberrant parameter values that signal model error.
3) We desire a parameter estimation algorithm that is easy to use, even for individual scientists who have expertise in cloud parameterization but not in parameter estimation.
4) Our source of "data" is LES output that is based on observed cases. Using LES output as data has two advantages: the LES model can be set up identically to the SCM, and the LES outputs difficult-to-observe fields such as higher-order moments, liquid water, and cloud fraction. Our goal is limited to emulating LES output, not observational data. Therefore, we treat the LES data as perfect input. Improving the agreement between LES and observations is a separate project that is beyond our scope.

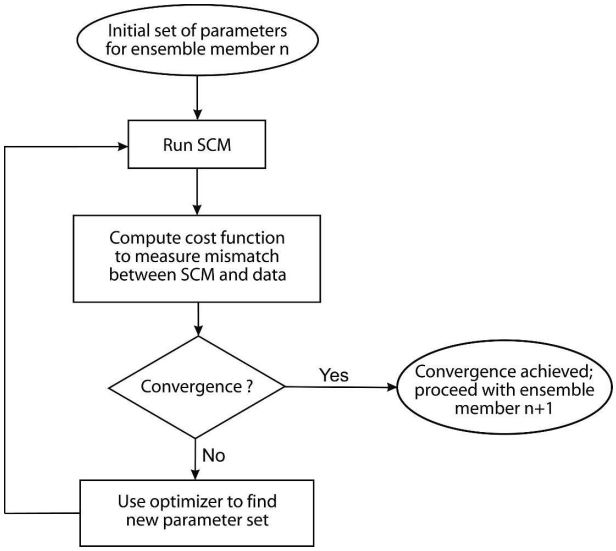

FIG. 1. Flowchart illustrating the optimization algorithm for a single ensemble member.

Our parameter estimation algorithm allows complete flexibility in the choice of field(s) to be optimized and parameter(s) to be estimated. Namely, we can optimize any combination of prognostic or diagnostic fields produced by the SCM and contained in the LES data. We can do so over any time period(s) or altitude range. Furthermore, we can estimate any combination of parameters in the SCM. We can also find overall best-fit parameters for two or more cases (e.g., cumulus and stratocumulus) at once.

Our parameter estimation algorithm for a single ensemble member is as follows. Before beginning the parameter estimation procedure, we select the SCM and LES output fields that we wish to match. Then we decide which SCM parameters to calibrate, and we choose initial values for these parameters. Then we perform the following steps (see Fig. 1 for a flowchart): run the SCM and evaluate the mismatch between the SCM and LES using a cost function. If the mismatch falls below a predetermined threshold, the algorithm stops. Otherwise, the optimizer chooses a new set of parameter values and the procedure is repeated until convergence. The same procedure is repeated for each ensemble member but with different initial parameter values.

Central to the parameter estimation algorithm is the choice of the cost function, J. When errors in the data are assumed to be Gaussian, the cost function takes the generic form:

$$J(\mathbf{m}) = \sum_{i=1}^{N} \frac{1}{2N} \{[\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}]^{\text{T}} \mathbf{C}^{-1} [\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}]^{\text{T}}\}_i, \tag{1}$$

where N is the number of observations sets. For each set, there are M observations represented by the vector $\mathbf{d}_{\text{obs}}$. Here $\mathbf{g}(\mathbf{m})$ is the corresponding vector of model predictions obtained with the model parameter set $\mathbf{m}$ and $\mathbf{C}^{-1}$ is the inverse of the covariance matrix. It is often simplified by only keeping its diagonal elements, namely, the inverse of the squared variances of the observations (e.g., Jackson et al. 2003, 2004). We apply the same simplification here.

In our application, $\mathbf{g}(\mathbf{m})$ is obtained from the SCM output [denoted by $\mathbf{g}_{\text{SCM}}(\mathbf{m})$] and $\mathbf{d}_{\text{obs}}$ from the LES "observations" ($\mathbf{d}_{\text{LES}}$). Both $\mathbf{g}_{\text{SCM}}(\mathbf{m})$ and $\mathbf{d}_{\text{LES}}$ can comprise any combination of variables produced by both the SCM and the LES. They could include mean profiles, such as liquid water potential temperature $\overline{\theta}_l$, total water mixing ratio $\overline{q}_t$, cloud fraction, or cloud water mixing ratio $\overline{q}_c$. They could also include any one of the vertical turbulence moments, such as $\overline{w'^2}$ or $\overline{w'^3}$. An arbitrary number of variables can be included in the cost function. The observations can also include an arbitrary number of LES cases (e.g., cumulus and stratocumulus). Our reference LES data consist of 1-min averages. Because the LES fields fluctuate intermittently, we do not desire the SCM to mimic a particular LES evolution minute by minute, but rather the LES evolution averaged over a longer time window, typically about 1 h. The specific form of the cost function becomes

$$\begin{aligned} J(\mathbf{m}) &= \sum_{i=1}^{N_c} \sum_{j=1}^{N_v} \frac{1}{\sigma_{i,j}^2} \sum_{t_n} \left\{ \left[\overline{\mathbf{g}_{\text{SCM}}(\mathbf{m}) - \overline{\mathbf{d}_{\text{LES}}}^{t_n}} \right]^{\text{T}} \right. \\ &\quad \left. \times \left[\overline{\mathbf{g}_{\text{SCM}}(\mathbf{m}) - \overline{\mathbf{d}_{\text{LES}}}^{t_n}} \right] \right\}_{(i,j)} \\ &\equiv \sum_{i=1}^{N_c} \sum_{j=1}^{N_v} \frac{1}{\sigma_{i,j}^2} S_{i,j}(\mathbf{m}). \end{aligned} \tag{2}$$

Summations are performed over a number of cases (N_c), variables (N_v), and time windows (t_n). The overbar $\overline{(\,)}^{t_n}$ denotes the time-averaging operator applied to a specific time window. Both $\mathbf{g}_{\text{SCM}}$ and $\mathbf{d}_{\text{LES}}$ are vectors that contain LES and SCM data from every vertical level within the altitude range of interest. Here $S_{i,j}(\mathbf{m})$ is the contribution to the error from a combination of a particular case (i) and variable (j).

Because we treat the LES observations as perfect inputs for the purpose of calibrating our SCM parameters, we do not have uncertainties attached to our observations that can be used to estimate the desired variances $\sigma_{i,j}$. Instead, we set $\sigma_{i,j}$ in Eq. (2) such that contributions from each case and each variable have equal weights in the initial value of the cost function:

$$\sigma_{i,j}^2 = S_{i,j}(\mathbf{m}_0) \quad \text{for} \quad i = 1 \ldots N_c \quad \text{and} \quad j = 1 \ldots N_v, \tag{3}$$

where $S_{i,j}(\mathbf{m}_0)$ denotes the *initial* error contribution. Equation (3) allows us to include disparate variables and results in each dataset and variable having an equal opportunity at reducing the overall cost function. Note that this initial weighting could be altered if a modeler had expert opinion indicating that one case or one variable should be overweighed compared to another one.

The optimization algorithm that we use is the downhill simplex method (Press et al. 1992). The simplex algorithm is not as efficient as some others, but high efficiency is not our first priority because SCMs are computationally inexpensive. For instance, our SCM can simulate 6 h of cloud evolution within seconds on a desktop personal computer. An additional benefit of the method is that it avoids the need to develop a model adjoint code, thereby greatly simplifying the coding and implementation. We stress that the choice of a particular optimization scheme is not of fundamental importance. The SCM's low computational expense allows us to employ other readily available optimization schemes, such as a conjugate gradient method with finite-difference estimates of the Jacobian.

The minimization of J proceeds on a N-dimensional surface, where N is the number of parameters that we wish to estimate. Because of the complexity and dimensionality of J, the topology likely consists of a large number of local valleys and floors where the minimization may stop. Therefore, it would be naive to assume that a single optimization will reach the best minimum. In fact, a single "best" minimum may not even exist; instead, a given function J may possess many comparably good minima. For this reason, we choose to perform an *ensemble* of minimizations.

Each ensemble member starts from slightly different initial conditions. The initial conditions consist of a simplex of $N + 1$ points on the N-dimensional surface. For instance, on a two-dimensional surface, the initial simplex is a triangle. The initialization is schematically illustrated in Fig. 2 for a two-dimensional case with four ensemble members. For each parameter to be cali-

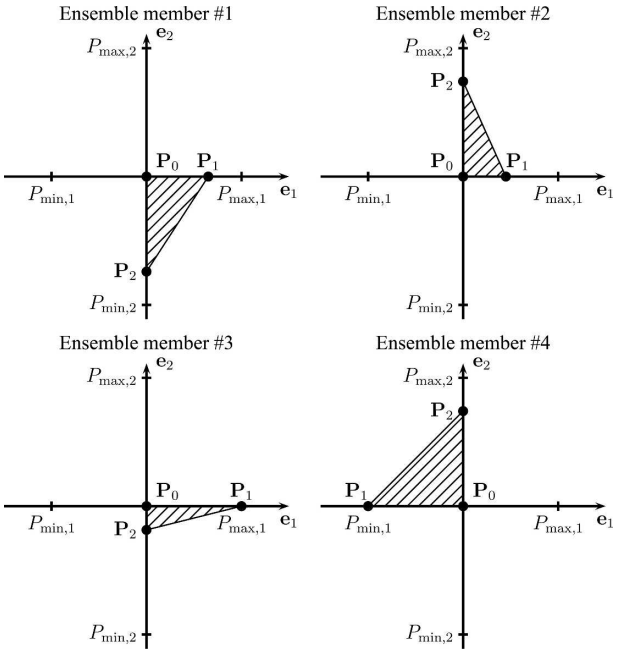

FIG. 2. Illustration of the ensemble initialization procedure for a two-dimensional problem. The hatched triangles represent the initial simplex for four different ensemble members. They are all centered around the same point P_0. Points P_1 and P_2 are randomly perturbed along the first and second dimension, respectively, such that they fall within the allowable initial range ($P_{\min,j}$, $P_{\max,j}$).

brated, a broad range of values ($P_{\min,i}$, $P_{\max,i}$) is chosen. The origin of the simplex (P_0) lies in the middle of this parameter space while the remaining vertices are randomly chosen by widely perturbing P_0 along one particular dimension. This procedure is designed to assume little about the parameter values or distributions a priori; that is, we assume an approximately uniform prior parameter distribution. Furthermore, the range ($P_{\min,i}$, $P_{\max,i}$) applies only to the initial conditions; the optimization algorithm is free to wander outside this range.

Since each ensemble member starts from a slightly different initial simplex, it yields a different optimized parameter set. This ensemble approach would be wasteful if the model structure were perfect and the topology of cost function simple: then each ensemble member would produce identical results. However, the complexity of the SCM creates complex structures in the cost function topology. It is possible to have different parameter sets that yield similar cost function values. As a result, an ensemble approach is well suited because it can more fully explore the cost function space.

Because a single SCM simulation takes on the order of tens of seconds, the overall computational cost of the methodology is acceptable. An optimization for one initial set of parameters requires on the order of $O(100)$ iterations before converging. Furthermore, each optimization within an ensemble is independent of the others, thus making the methodology an "embarrassingly parallel problem" up to the ensemble size.

TABLE 1. Initial values and ranges of the 10 parameters for the initial calibration experiments. Also shown are the model variables or components directly affected by these parameters.

Parameter	Direct impact	$P_{0,j}$	Initial range $P_{\min,j}$	Initial range $P_{\max,j}$
C_1	$\overline{w'^2}$ (dissipation)	2.5	1.0	4.0
C_2	$\overline{q_t'^2}$, $\overline{\theta_l'^2}$, $\overline{q_t'\theta_l'}$ (dissipation)	1.5	0.5	2.5
C_5	$\overline{w'^2}$ (pressure term)	0.3	0.1	0.5
C_6	$\overline{w'q_t'}$, $\overline{w'\theta_l'}$ (pressure terms)	7.5	3.0	12.0
C_7	$\overline{w'q_t'}$, $\overline{w'\theta_l'}$ (pressure terms)	0.5	0.1	0.9
C_8	$\overline{w'^3}$ (pressure term)	2.5	1.0	4.0
C_{11}	$\overline{w'^3}$ (pressure term)	0.5	0.1	0.9
β	PDF functional form	1.25	0.5	2.0
$\tilde{\sigma}_w^2$	PDF functional form	0.35	0.3	0.4
μ ($\times 10^{-4}$ s^{-1})	Mixing length	6.0	4.0	8.0

Our approach to parameter estimation can be cast as an approximation to a Bayesian stochastic inversion with a uniform prior parameter distribution (Jackson et al. 2004). Each ensemble member of optimized parameter values does not represent a random draw from the correct posterior distribution, but rather needs to be weighted by its (possibly scaled) likelihood given the LES data. The scaling is necessary to account for inaccurate estimates of data uncertainty. We approximate this scaled weighting below by using the 20 ensemble members with the highest likelihood (lowest cost function value). This represents a suboptimal weighting that will produce biases in the estimates of the posterior distribution, but we feel that the inaccuracy is unimportant for our qualitative application.

4. Initial parameter estimation experiments

a. Configuration

A total of 10 parameters from the SCM have been selected for the initial calibration: C_1, C_2, C_5, C_6, C_7, C_8, C_{11}, β, $\tilde{\sigma}_w^2$, and μ. The initial mean values of the 10 parameters and their allowable range at the initial time are listed in Table 1. The actual model equations in which all these parameters occur can be found in Golaz et al. (2002a) and Larson and Golaz (2005). For convenience, the prognostic equations are also listed in the appendix. The C_1 controls the dissipation rate of the vertical velocity variance $\overline{w'^2}$. The C_2 controls the dis-

sipation rates of the scalar variances and covariance $\overline{q_t'^2}$, $\overline{\theta_l'^2}$, and $\overline{q_t'\theta_l'}$. The C_5 appears in the parameterization of the pressure correlation term in the $\overline{w'^2}$ equation. Both C_6 and C_7 appear in the pressure correlation terms of the scalar flux equations $\overline{w'q'_t}$ and $\overline{w'\theta_l'}$. Both C_8 and C_{11} are part of the pressure term parameterization in the third moment vertical velocity $\overline{w'^3}$. The parameters β and $\tilde{\sigma}_w^2$ arise from the PDF functional form. The β appears in the diagnostic relationship linking the skewness of θ_l and q_t to the predicted skewness of w. The $\tilde{\sigma}_w^2$ controls the width of the individual Gaussians in the PDF. Finally, μ is a mixing time scale used in the computation of the mixing length.

Initially, we estimate parameter values for two boundary layer cloud regimes separately. The differences in parameter values help reveal model structural errors. Then we estimate parameter values for both cases simultaneously. Both cases were previously simulated in LES intercomparisons of the Global Water and Energy Experiment (GEWEX) Cloud System Studies Working Group 1 (GCSS-WG1). The setup of both cases is based on observations. The first regime is a trade wind cumulus regime based on the Barbados Oceanographic and Meteorological Experiment (BOMEX; Siebesma et al. 2003). The second is a marine stratocumulus case, DYCOMS-II RF01, hereafter referred to as RF01 (Stevens et al. 2005). For each case, the SCM is calibrated against LES results obtained with a version of the Coupled Ocean–Atmosphere Mesoscale Prediction System (COAMPS) that is suitably modified for LES scales, which we call "COAMPS–LES" (Golaz et al. 2005). The COAMPS–LES results look similar to other LES results in the intercomparisons. BOMEX and RF01 are selected because they represent different ends of the boundary layer cloud regime spectrum. We calibrate only two cases in order to avoid overfitting.

The variables appearing in the cost function in (2) are chosen to be cloud fraction and cloud water mixing ratio. Because a major goal of the SCM is to produce realistic clouds, the selection of these cloud variables is natural. The time windows t_n are as follows. BOMEX is a 6-h-long simulation, and we select three 1-h windows, consisting of the fourth, fifth, and sixth simulation hours. RF01 is a 4-h-long simulation, and we select the third and fourth hours as time windows. The initial experiments we present consist of three ensembles: one that uses BOMEX data exclusively (B1) in the optimization, a second that uses only RF01 data (D1), and a third that combines both BOMEX and RF01 (BD1). Each experiment consists of an ensemble of 400 members. A complete list of all experiments is provided in Table 2.

TABLE 2. Summary of all ensemble-based parameter estimation experiments. Cloud fraction is abbreviated as *c*.

Category	Name	Size	No. of parameters	LES cases	LES fields
Initial	B1	400	10	BOMEX	c, $\overline{q}_c$
	D1	400	10	RF01	c, $\overline{q}_c$
	BD1	400	10	BOMEX and RF01	c, $\overline{q}_c$
Revised	B2	400	14	BOMEX	c, $\overline{q}_c$
	D2	400	14	RF01	c, $\overline{q}_c$
	BD2	400	14	BOMEX and RF01	c, $\overline{q}_c$

b. Results

Results of the initial parameter estimation experiments are shown as scatterplots in Fig. 3. In the scatterplots, each dot represents one ensemble member. The dots are color coded by experiment: green for BOMEX (B1), red for RF01 (D1), and blue for the combined experiment (BD1). The horizontal axes represent the final parameter value, and the vertical axes represent the normalized cost function end value: $\hat{J} = J/J_{\min}$, where $J_{\min}$ is the lowest cost function value of the ensemble. Here $J_{\min}$ is computed separately for each ensemble. Therefore, the best fitting members (as measured by J) within a given ensemble reside on the lower portion of each panel, and the worst reside in the upper portion.

For each of the 10 parameters, the scatterplots reveal a surprisingly large spread in the final parameter values compared to the initial range (gray shaded area). The plots clearly illustrate the implausibility of finding a global minimum that is substantially better than other local minima and justifies the use of an *ensemble-based* optimization approach. For a number of parameters, the posterior parameter distribution is actually wider than the prior distribution. This seems counterintuitive at first, but it results from the fact that parameter values covary with each other. The product of the optimization is really an ensemble of *parameter sets* drawn from a single 10-dimensional distribution and not independent parameters drawn from 10 separate 1D distributions. Scatterplots can only depict the marginal projections of this multidimensional distribution and cannot reveal how parameters covary. Therefore, changing one and only one parameter value to another arbitrary value within the range of the scatterplot is likely to worsen the fit, because it would neglect the covariation with other parameters. Also, because of this covariation between parameters, it would not be justifiable to select the mean of each marginal parameter distribution as an optimum parameter value. Identification of the covari-

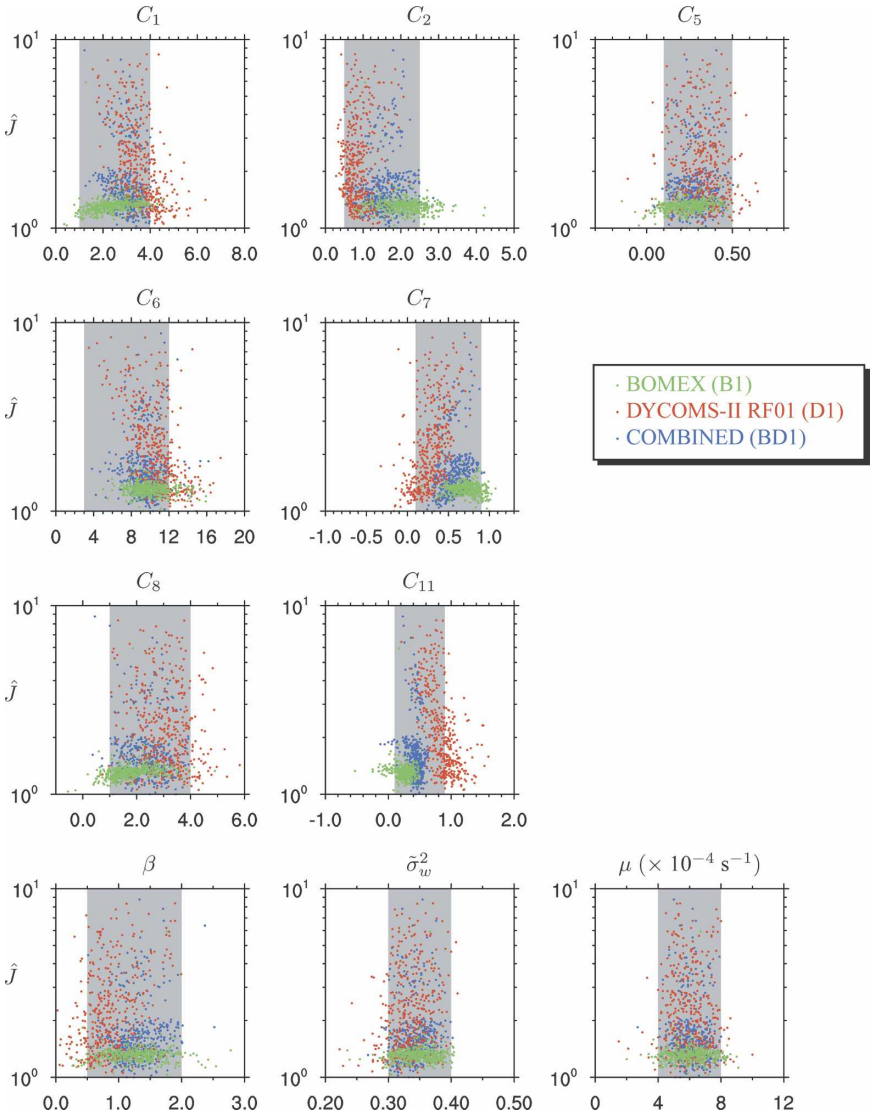

FIG. 3. Results of the initial parameter estimation experiments (B1, D1, and BD1). Each panel represents 1 of the 10 parameters. The horizontal axis is the final parameter value, and the vertical axis the normalized error, $\hat{J}$, of a given optimization with respect to the best member of the ensemble. Each optimization is represented by a single dot. Green dots are for the BOMEX ensemble (B1), red dots are for the RF01 ensemble (D1), and blue dots are for the combined ensemble (BD1). The gray shaded areas indicate the initial allowable parameter ranges.

ance between parameters is a benefit of the ensemble approach that is not exploited in the current work. Understanding how best to utilize the covarying information is the subject of ongoing research.

It is also interesting to note that some parameter values lie outside their physically expected range. Specifically, C_5, C_7, C_8, and C_{11} are expected to be positive but are negative in some optimization runs. Our algorithm does not explicitly restrict the range of parameter values, except when the values numerically destabilize the SCM, in which case the cost function is set to a large penalty value.

The optimized parameter distribution reveals some unexpected features. For some parameters, the distributions for BOMEX (green dots) and RF01 (red dots) overlap considerably, whereas other parameter distributions overlap only slightly. In particular, note the small overlap between green and red dots for C_7 and C_{11}. This small overlap indicates underfitting, which is symptomatic of model structural error.

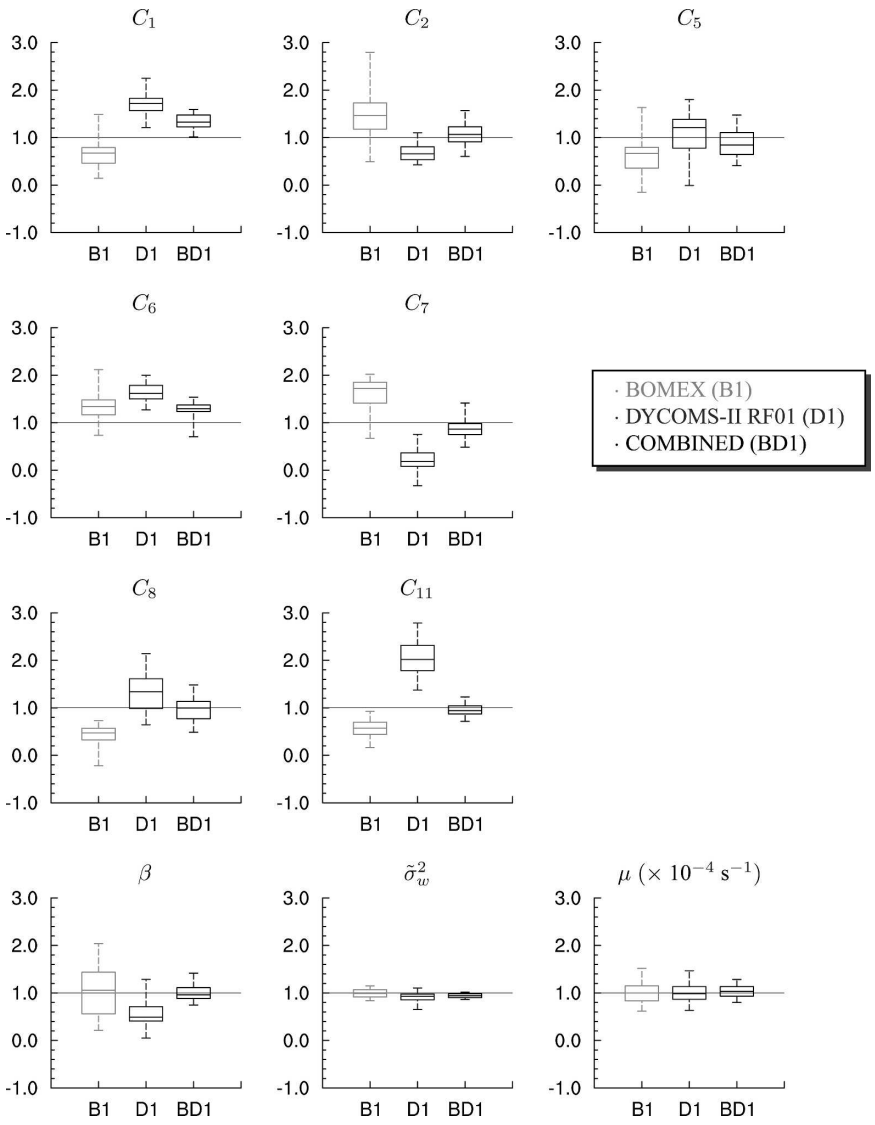

FIG. 4. Distributions of the final parameter values for the 20 lowest J ensemble members of the initial parameter estimation experiments (B1, D1, and BD1). The values are normalized by the initial parameter value to show the departure of the end value compared to the initial value. Each box plot shows the minimum, first quartile, median, third quartile, and maximum values of the distribution.

Figure 4 displays the same results using standard box plots. Because the quality of the parameter sets can vary substantially between the best and the worst ensemble member, we focus only on the 20 best ensemble members for each experiment. The parameters are normalized by their initial values to show their relative departure. These box plots clearly reveal that the calibration experiments for BOMEX and RF01 tend to favor different values for C_7 and C_{11}. Parameter values for the combined experiment lie in between.

Profiles from the SCM simulations using the 20 best parameter sets are depicted in Figs. 5 and 6. The profiles shown are mean liquid water potential temperature ($\overline{\theta}_l$), mean total and cloud water mixing ratios ($\overline{q}_t$, $\overline{q}_c$), cloud fraction, and the second and third central moments of the vertical velocity ($\overline{w'^2}$, $\overline{w'^3}$). For BOMEX (Fig. 5), the calibrated SCM runs are able to adequately reproduce the LES profiles. Note that only the cloud fraction and $\overline{q}_c$ enter the definition of the cost function J. The $\overline{\theta}_l$, $\overline{q}_t$, $\overline{w'^2}$, $\overline{w'^3}$ are not directly driven to match the corresponding LES profiles. This shows that reasonable physical constraints are embedded in the SCM. However, none of the simulations accurately reproduces the cloud fraction near the cloud base. This ap-

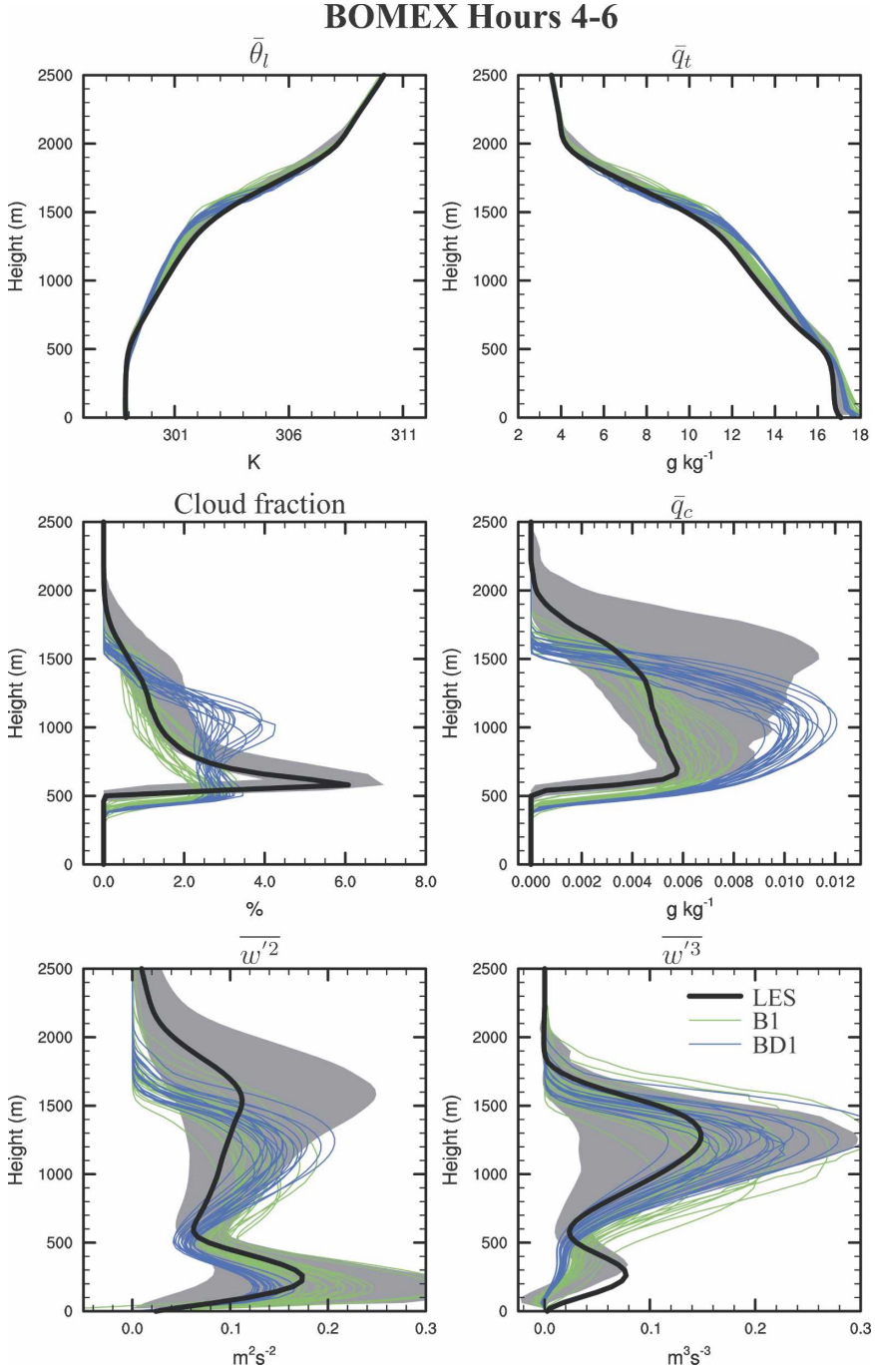

FIG. 5. Comparison of the COAMPS–LES profiles (black) with the 20 lowest J value SCM simulations for the BOMEX ensemble (B1, green) and combined ensemble (BD1, blue) of the initial parameter estimation experiments. Profiles shown are liquid water potential temperature ($\bar{\theta}_l$), total and cloud water mixing ratios ($\bar{q}_t$, $\bar{q}_c$), cloud fraction, and second and third moments of the vertical velocity ($\overline{w'^2}$, $\overline{w'^3}$). They are averaged over the last 3 h of the simulation. The gray-shaded areas indicate the range (minimum and maximum bounds) of other LES models from the intercomparison.

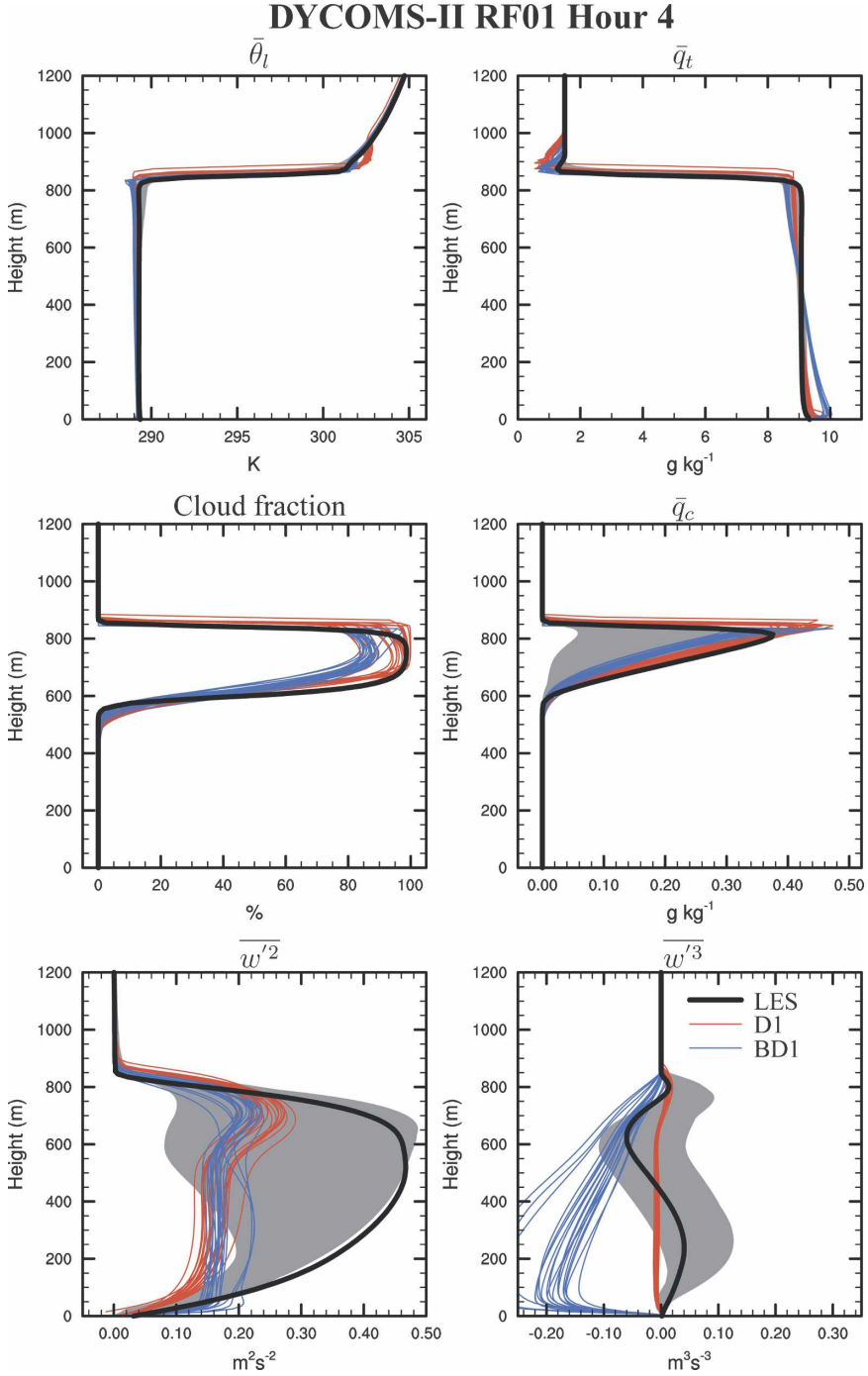

FIG. 6. Same as in Fig. 5, but for RF01. Red lines are SCM results from the RF01 ensemble (D1) and blue lines are from the combined ensemble (BD1). Profiles are averaged over the last simulation hour. LES model range is shown with gray-shaded areas, where available.

pears to be a manifestation of a model deficiency that cannot be corrected by simple parameter recalibration. The differences between the SCM members obtained from the BOMEX experiment (B1, green) and the combined one (BD1, blue) show larger values of liquid water and to a lesser degree cloud fraction for the combined experiments while the other profiles remain comparable.

The RF01 profiles paint a different picture (Fig. 6). Even though cloud water for the DYCOMS RF01 en-

semble (D1, red) and the combined ensemble (BD1, blue) appear comparable, some significant differences are present in other fields. In particular, the combined ensemble has difficulties reproducing the profiles of $\overline{q}_t$ and $\overline{w'^3}$. The SCM is unable to produce a well-mixed total water profile, an indication of a poor representation of boundary layer mixing processes. The $\overline{w'^3}$ is unrealistically negative in the lower portion of the domain.

The differences between C_7 and C_{11} values in BOMEX and RF01 warrant further discussion. The C_7 appears in the pressure correlation term for the turbulence fluxes $\overline{w'q'_t}$ and $\overline{w'\theta'_l}$:

$$\left.\frac{\partial \overline{w'q'_t}}{\partial t}\right|_{\text{pressure}} \equiv -\frac{1}{\rho_0}\overline{q'_t\frac{\partial p'}{\partial z}} = -\frac{C_6}{\tau}\overline{w'q'_t} - C_7\left(-\overline{w'q'_t}\frac{\partial \overline{w}}{\partial z} + \frac{g}{\theta_0}\overline{q'_t\theta'_v}\right), \tag{4}$$

$$\left.\frac{\partial \overline{w'\theta'_l}}{\partial t}\right|_{\text{pressure}} \equiv -\frac{1}{\rho_0}\overline{\theta'_l\frac{\partial p'}{\partial z}} = -\frac{C_6}{\tau}\overline{w'\theta'_l} - C_7\left(-\overline{w'\theta'_l}\frac{\partial \overline{w}}{\partial z} + \frac{g}{\theta_0}\overline{\theta'_l\theta'_v}\right). \tag{5}$$

The C_{11} enters in the parameterization of the pressure correlation term of $\overline{w'^3}$:

$$\left.\frac{\partial \overline{w'^3}}{\partial t}\right|_{\text{pressure}} \equiv -\frac{3}{\rho_0}\overline{w'^2\frac{\partial p'}{\partial z}} = -\frac{C_8}{\tau}(C_{8b}\text{Sk}_w^4 + 1)\overline{w'^3} - C_{11}\left(-3\overline{w'^3}\frac{\partial \overline{w}}{\partial z} + \frac{3g}{\theta_0}\overline{w'^2\theta'_v}\right). \tag{6}$$

Each of the pressure correlation parameterizations above contains three terms: the first term is a Newtonian damping term, the second term is proportional to $\partial\overline{w}/\partial z$ and is generally negligible, and the third term is proportional to the relevant buoyancy moment. The Newtonian term is sometimes called a "return to isotropy" or "slow" term. It always has the opposite sign of the prognosed moment and therefore acts as a sink. The last two terms are sometimes called "rapid" terms (Pope 2000). The buoyancy term can act as a sink or a source depending on the sign of the buoyancy moment.

LES data reveal that $\overline{w'^3}$ and $\overline{w'^2\theta'_v}$ usually have the same sign, making C_{11} an additional damping term. Cumulus layers have large skewness values, and these must be permitted by the SCM in order to obtain reasonable cloud properties. To accomplish this, the BOMEX calibration experiment reveals that the SCM needs reduced pressure damping and hence smaller C_{11} values. For stratocumulus, the skewness tends to be small due to the near symmetry between updraft and downdraft velocities. Maintaining low skewness in RF01 is aided by larger C_{11} values.

The role of the C_7 term is more complicated. The buoyancy contribution to the pressure correlation term in Eqs. (4) and (5) is largest in the upper portion of the cloud layer, both for BOMEX and RF01, near an inversion. The inversion is weak and several hundred meters deep for BOMEX, but very sharp and shallow for RF01. Physically, rising eddies are impeded in their ascent by the presence of the inversion. The kinetic energy of impeded updrafts is redistributed from the vertical to the horizontal by the pressure correlation terms. As a result, the buoyancy contribution to the pressure term strongly affects the turbulence fluxes $\overline{w'q'_t}$ and $\overline{w'\theta'_l}$ near the inversion and thereby cloud-top mixing. As shown by LESs, RF01 is particularly sensitive to cloud-top mixing and entrainment (Stevens et al. 2005). The tight coupling between cloud-top mixing and C_7 is the likely cause of RF01's preference for lower values of C_7, as revealed by the calibrations.

c. Summary

The initial calibration experiments have revealed that the SCM simulations agree relatively well with the reference LES for both BOMEX and RF01, if the SCM uses separately calibrated parameter values. A reasonable time evolution of RF01 can also be simulated without having to change the mixing length formulation, a finding that was not at all obvious before the experiments were performed.

However, the experiments clearly demonstrate that improved BOMEX and RF01 require different values of the parameters C_7 and C_{11}. This is undesirable since the SCM is intended to serve as a general boundary layer parameterization. We address this issue further in the next section.

5. Revised parameter estimation experiments

a. Proposed model modifications

One major difference between BOMEX and RF01, or more generally between shallow cumulus and stratocumulus clouds, is the vertical velocity skewness, $\text{Sk}_w = \overline{w'^3}/\overline{w'^2}^{3/2}$. The skewness of w measures the

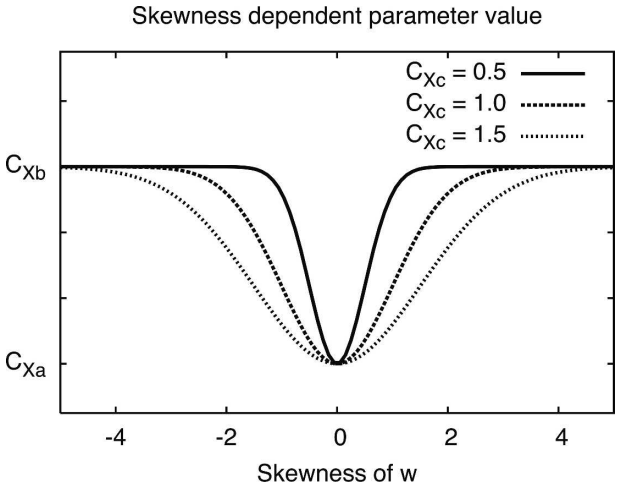

FIG. 7. Illustration of the skewness-dependency function of the modified parameters in Eqs. (7)–(8). Different values of the denominator in the exponent are plotted for illustration purposes.

asymmetry between updrafts and downdrafts. In shallow convection, updrafts tend to be narrow and strong and the compensating downdrafts are broad and weak, giving rise to a large positive skewness. For stratocumulus, areas and vertical velocities of updrafts and downdrafts tend to be comparable, which translates into small positive or negative skewness values. Based on the findings of the previous section, we propose to reformulate the parameters C_7 and C_{11} so as to convert them into skewness-dependent functions:

$$C_7(\mathrm{Sk}_w) = C_{7b} + (C_{7a} - C_{7b})e^{-1/2(\mathrm{Sk}_w/C_{7c})^2}, \quad (7)$$

$$C_{11}(\mathrm{Sk}_w) = C_{11b} + (C_{11a} - C_{11b})e^{-1/2(\mathrm{Sk}_w/C_{11c})^2}. \quad (8)$$

Equation (7) implies that in the limit of small skewness magnitudes, $C_7 \rightarrow C_{7a}$, and in the limit of large skewness, $C_7 \rightarrow C_{7b}$. The sharpness of the transition between small and large values is controlled by C_{7c}. The form of the skewness dependency is illustrated in Fig. 7. Equations (7)–(8) are purely empirical and we make no attempt to justify them theoretically. In essence, they are simple structural modifications that one can conceive to remedy the problems uncovered in the previous section.

A new set of ensemble-based parameter estimation experiments is performed using the new formulations for C_7 and C_{11} (B2, D2, and BD2 in Table 2). The methodology is the same as for the initial experiments, but the dimensionality of the optimization problem is now 14. The range and initial values of the newly introduced parameters are given in Table 3. All the other parameters have the same range and value as in Table 1.

TABLE 3. Initial values and ranges of the new parameters for the revised calibration experiments (B2, D2, and BD2). The initial values and ranges of the remaining parameters are identical to those in Table 1.

Parameter	$P_{0,j}$	Initial range $P_{\min,j}$	Initial range $P_{\max,j}$
C_{7a}	0.5	0.1	0.9
C_{7b}	0.5	0.1	0.9
C_{7c}	1.0	0.25	1.75
C_{11a}	0.5	0.1	0.9
C_{11b}	0.5	0.1	0.9
C_{11c}	1.0	0.25	1.75

b. Results

The final parameter values of all the members for the revised experiments B2, D2, and BD2 are shown as scatterplots in Fig. 8. The overlap between BOMEX (green points) and RF01 (red points) ensembles for C_{7x} and C_{11x} is now improved compared to Fig. 3. As a caveat, we note that because the SCM inevitably still contains structural errors and because we have optimized simultaneously for all parameter values, the optimized values have undoubtedly been influenced by compensating errors between terms. Therefore, an optimal parameter value for a given term in our SCM is not necessarily optimal for the same term in a different SCM.

We now focus on the SCM profiles of the 20 best parameter sets of each ensemble. The profiles from the BOMEX ensemble (B2, green lines; Fig. 9) show little change compared to the initial experiment (B1, green lines; Fig. 5). The cloud water profile is improved in the revised combined experiment (BD2, blue lines), but the cloud fraction is still underestimated near the cloud base.

The impact of the modified pressure terms C_7 and C_{11} is more significant for RF01 (Fig. 10). The RF01 ensemble (D2) and the combined ensemble (BD2) now both yield SCM results that agree better with COAMPS–LES. This is in contrast to the initial experiments (Fig. 6) in which the RF01 ensemble (D1) produced total water mixing ratio profiles that were not sufficiently well mixed, and had erroneous $\overline{w'^3}$ profiles.

The results from the BD2 ensemble demonstrate that the modifications made to C_7 and C_{11} in Eqs. (7)–(8) allow for the existence of parameter sets that produce reasonable results for BOMEX and RF01 *simultaneously.* This was not the case with the unmodified SCM. Furthermore, before this work was performed, it

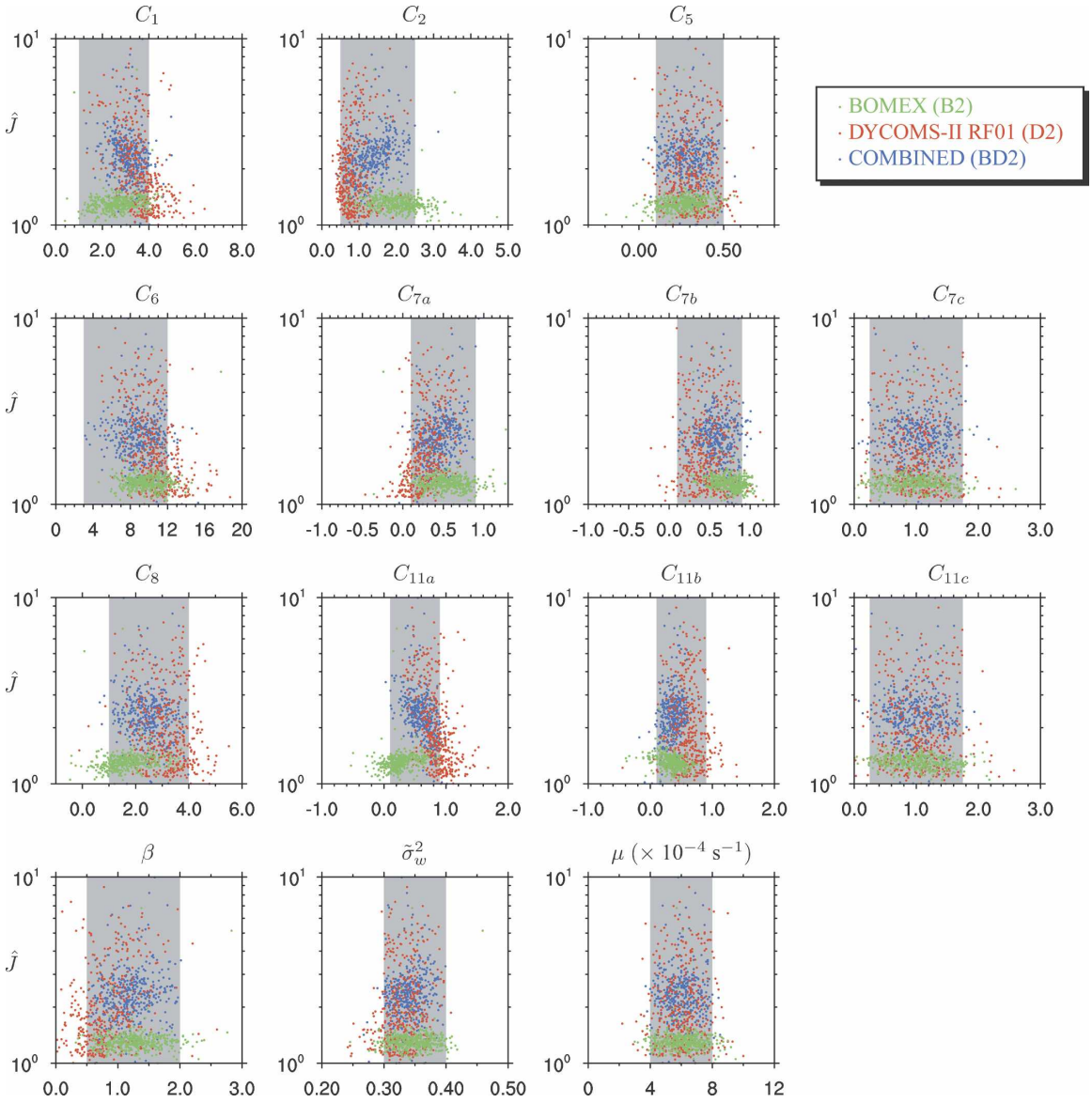

FIG. 8. Same as in Fig. 3, but for the results of the revised parameter estimation experiments (B2, D2, and BD2) with 14 parameters. Green dots are for the BOMEX ensemble (B2), red dots are for the RF01 ensemble (D2), and blue dots are for the combined ensemble (BD2).

would have been difficult to identify modifications to the SCM that would have been likely to faithfully simulate both BOMEX and RF01.

6. Evaluation with independent data

We have intentionally calibrated only two LES cases and reserved other cases for cross validation in order to avoid overfitting. To verify that we have indeed avoided overfitting, we simulate four additional test cases using the 20 best parameter sets from the BD1 and BD2 ensembles. The additional test cases are all set up according to the specifications of GCSS intercomparisons, which are based loosely on the observations.

The first case is shallow cumulus over land from the Southern Great Plains (SGP) Atmospheric and Radiation Measurement (ARM) site (Brown et al. 2002). The simulation starts in the morning with clear skies. As the day progresses, the boundary layer deepens due to the surface fluxes. Fair-weather cumulus clouds develop, grow progressively during the day, and dissipate before the simulation's end at about sundown.

The second case involves cumulus clouds rising under a broken stratocumulus deck (Stevens et al. 2001). It is

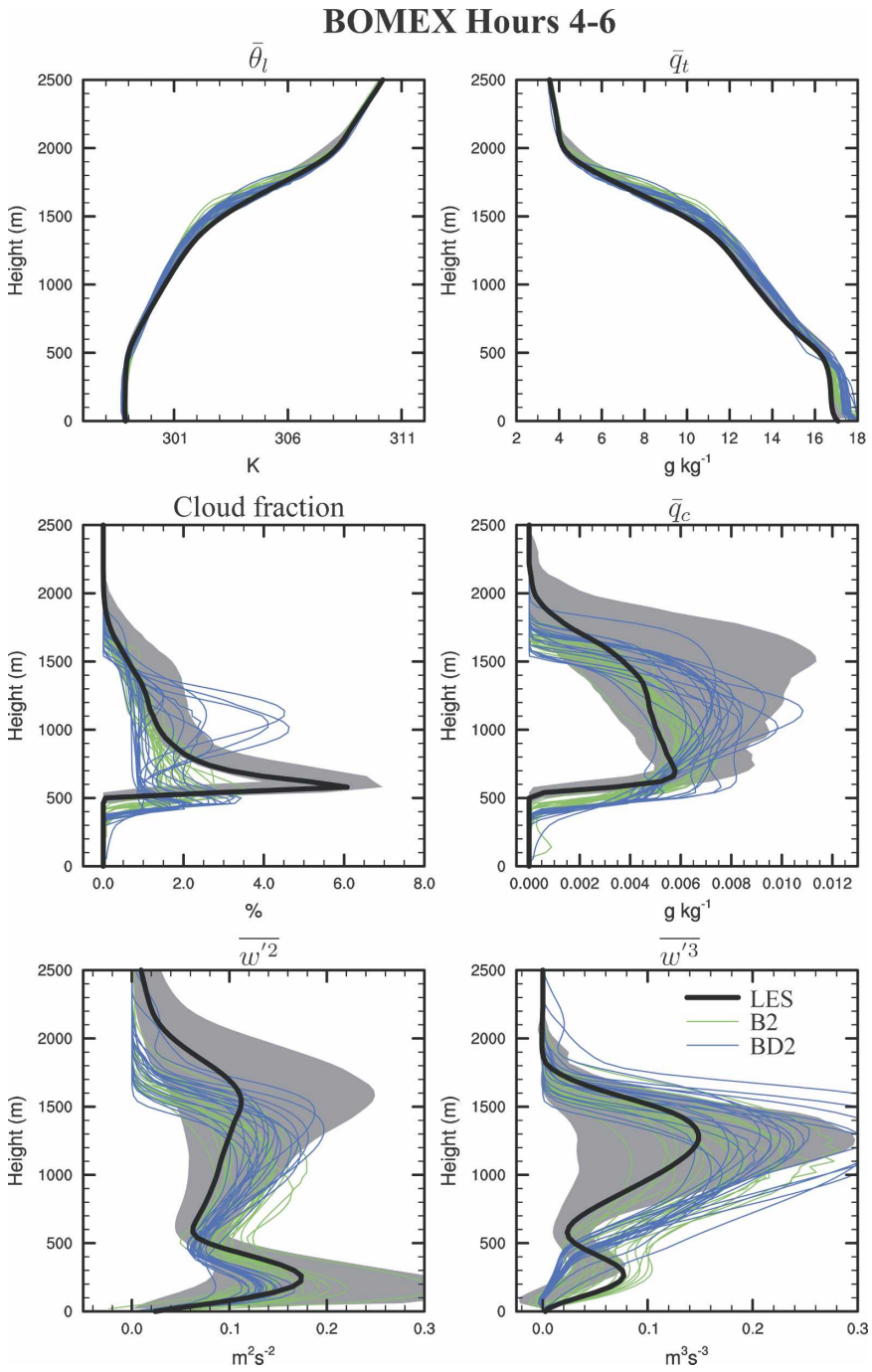

FIG. 9. Same as in Fig. 5, but for the revised parameter estimation experiments (B2 and BD2). Green lines are the SCM results from the BOMEX ensemble (B2) and blue lines are from the combined ensemble (BD2).

idealized from data collected during the Atlantic Trade Wind Experiment (ATEX). The cloud fraction in the cumulus layer is less than 10%. In the overlying stratiform layer, the cloud fraction varies significantly from one LES model to another; the range is approximately 20%–80%.

The last two cases are both nocturnal stratocumulus-topped layers. One is based on observations from the First International Satellite Cloud Climatology Project (ISCCP) Regional Experiment (FIRE; Moeng et al. 1996). There was a large spread in LES results, partly

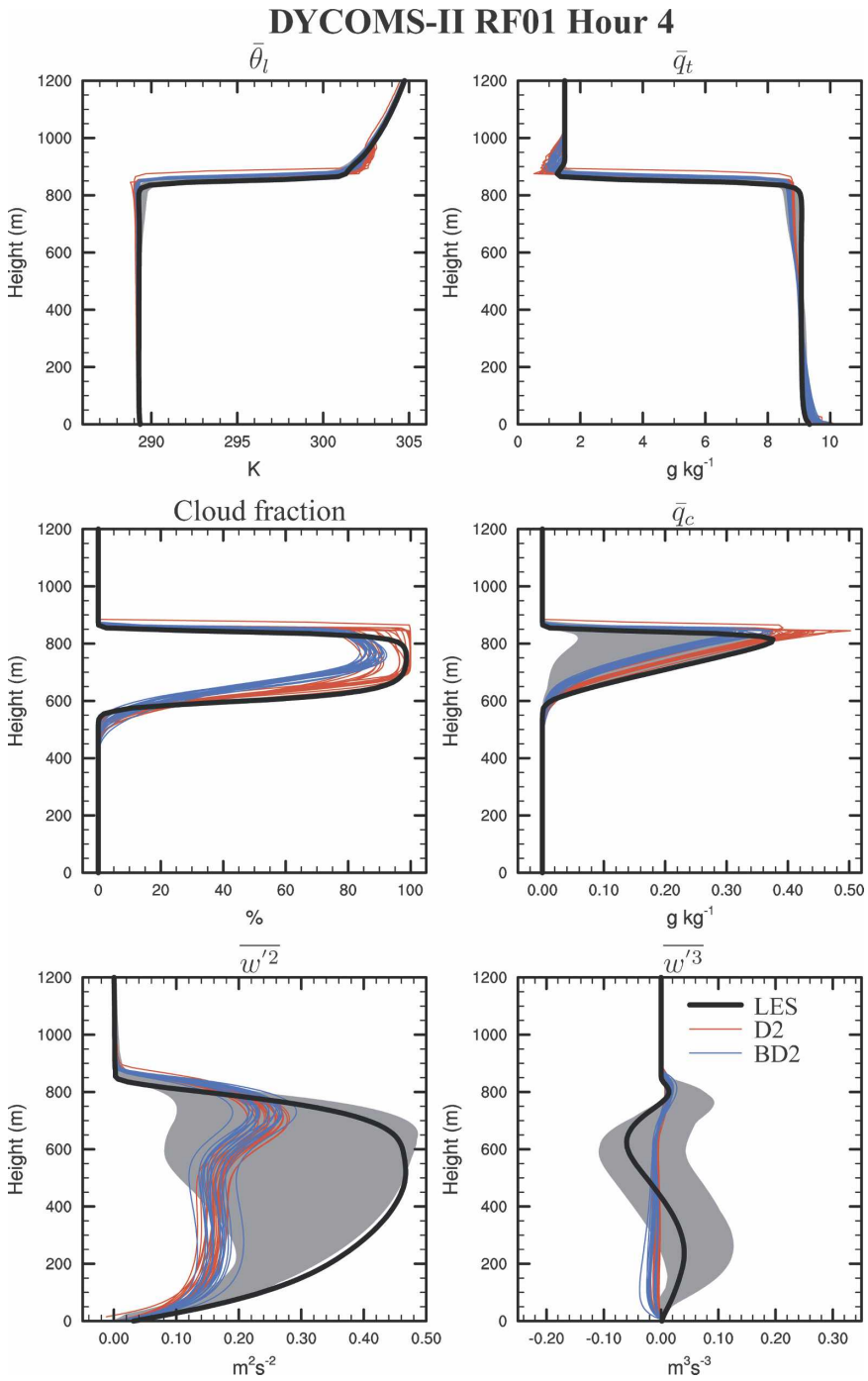

FIG. 10. Same as in Fig. 6, but for the revised parameter estimation experiments (D2 and BD2). Red lines are the SCM results from the RF01 ensemble (D2) and blue lines are from the combined ensemble (BD2).

because the intercomparison did not specify the form of the cloud-top radiative forcing. The second stratocumulus case is based on the second research flight (RF02) of the DYCOMS-II field experiment (for a description of the intercomparison, see http://sky.arc.nasa.gov:6996/ack/gcss9/index.html). Here we focus on the nondrizzling RF02 case.

Figure 11 shows the SCM cloud properties obtained with the 20 best parameter sets from the ensembles BD1 (blue) and BD2 (orange) and compares them with

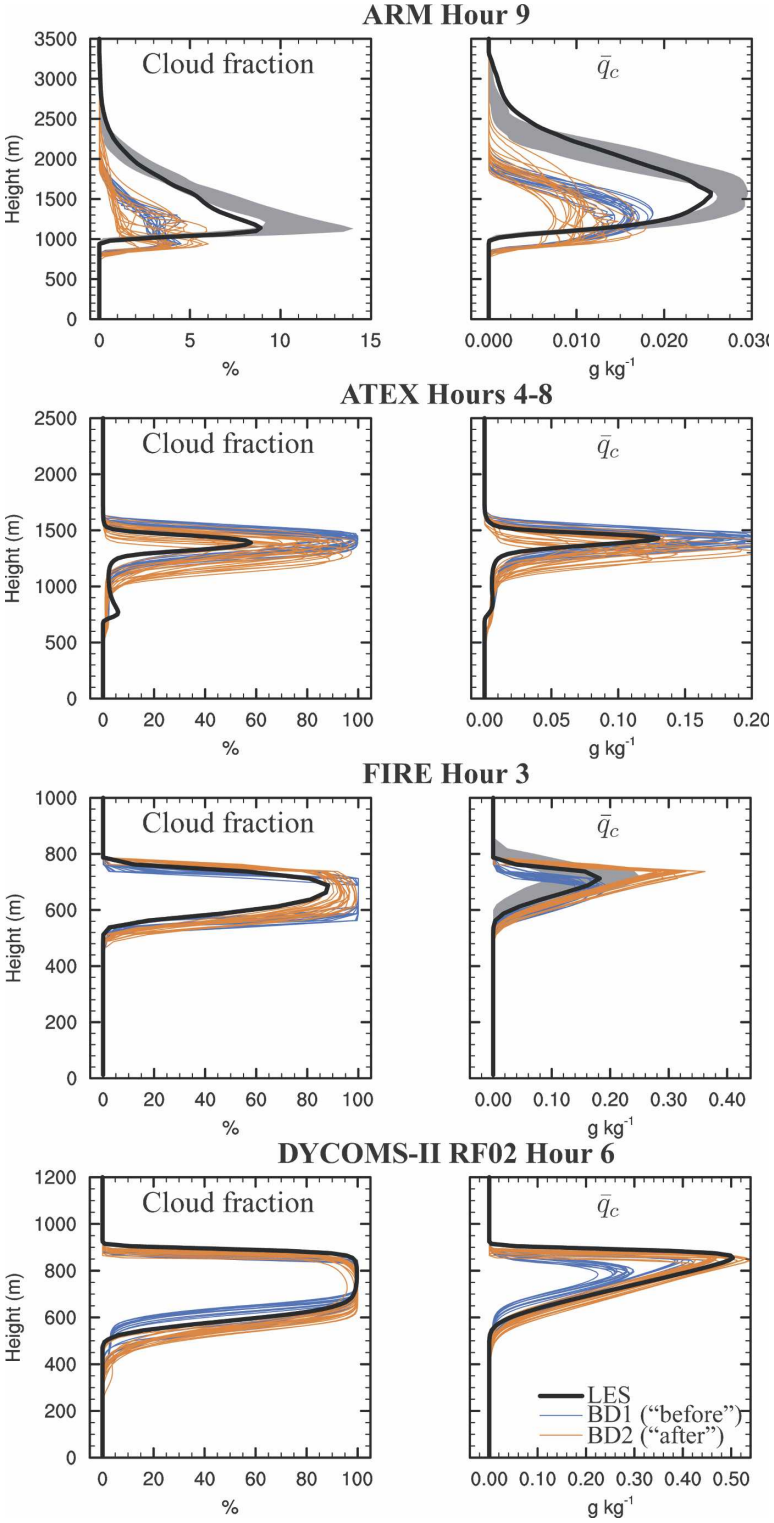

FIG. 11. Evaluation of the 20 best parameter sets obtained from the combined experiments BD1 and BD2 using independent datasets. Black lines are COAMPS–LES results and the gray-shaded areas represent LES ranges from model intercomparisons (where available). The SCM results from the 20 best parameter sets from BD1 (i.e., before empirical modifications) are plotted in blue. The BD2 (i.e., after empirical modifications) results are plotted in orange.

COAMPS–LES (black). The main interest is in comparing the BD1 and BD2 SCM profiles, which are the "before" and "after" pictures showing the effects of our empirical modifications.

The biggest difference occurs for RF02, where the BD2 SCM (after) cloud water profiles are markedly superior to the BD1 (before) profiles. Most BD1 ensemble members predict liquid water amount near the cloud top that underestimates the LES value by nearly 50%. In contrast, the BD2 ensemble members almost exactly match the LES.

The BD2 FIRE cloud properties appear to suffer a slight degradation in BD2 compared to BD1. Given the uncertainties surrounding the specifications of this case, one should be cautious with respect to the significance of this degradation. For example, a simulation using a different LES model produced a liquid water amount of 0.25 g kg^{-1} near the cloud top (Golaz et al. 2002b, their Fig. 12d), much closer to the BD2 values.

For ARM and ATEX, the changes between BD1 and BD2 are modest. ARM results both underestimate the cloud amount compared to the LES. These profiles are actually slightly worse than the results published in Golaz et al. (2002b). However, since BD1 and BD2 profiles are qualitatively similar, the degradation does not stem from the SCM modifications we made [Eqs. (7) and (8)], but rather from the values of the parameters themselves.

On balance, the empirical changes made to C_7 and C_{11} [Eqs. (7) and (8)] are beneficial to RF02, neutral to ARM and ATEX, and slightly negative for FIRE. Given that these changes are solely based on BOMEX and RF01 datasets, we can safely state that we have avoided an overfitting situation and hence can have some confidence in the generality of the SCM modifications, despite their empirical nature.

7. Conclusions

We have presented an ensemble method of parameter estimation. It has four chief advantages:

1) It allows complete flexibility in the choice of parameters to be estimated and fields to be optimized. For instance, we may simultaneously estimate any combination of the parameters in Tables 1 and 3 (i.e., C_1, C_2, and so forth). Furthermore, we may optimize any combination of fields (e.g., cloud fraction and liquid water) that are produced by the SCM and contained in the LES data.
2) The method is conceptually straightforward.
3) The method is easy to implement, because it does not require writing an adjoint of the model code.
4) The method produces an *ensemble* of sets of best-fit parameter values. This is useful in cases in which the cost function contains many comparable local minima. The ensemble methodology provides not only the range of acceptable values of parameters, but also information about how the best-fit parameter values covary with each other.

We have used the ensemble parameter estimation method to calibrate a single-column model (SCM) of boundary layer clouds. The "data" used are output from six large-eddy simulations (LES). These consist of three stratocumulus cases, a trade wind cumulus case, a continental cumulus case, and a cumulus-under-stratocumulus case. We calibrate 10 SCM parameters simultaneously against profiles of cloud fraction and liquid water.

In calibrating the SCM, we sought to avoid the opposing problems of overfitting and underfitting.

To avoid overfitting, we fit only two fields, cloud fraction and liquid water, and two cases, the BOMEX trade wind cumulus case and the DYCOMS-II RF01 marine stratocumulus case. Other fields and cases were used for cross validation. That is, they were used to verify that the chosen parameter values fit well generally, not merely for the two fields and cases used in the calibration.

To diagnose the cause of underfitting, we calibrated BOMEX and RF01 separately, thereby obtaining two sets of parameter values. The separate calibrations revealed differences in the values of the parameters C_7 and C_{11}. Assessing the significance of these differences was made possible by the ensemble methodology, which clearly showed the lack of overlap in the acceptable parameter values. This demonstrates that calibration need not obscure model structural error; in fact, if used strategically, calibration may *reveal* structural errors. We then replaced the parameters C_7 and C_{11} by the empirical functions of skewness. This structural modification ameliorated the underfitting and permitted the SCM to model all six cloud cases more accurately without case-specific adjustments.

Although the parameter estimation technique can help identify the existence of model structural errors, it cannot propose new ideas to fix those errors. Nevertheless, automated parameter estimation does speed up the process of model development because it allows rapid recalibration when a new model improvement is introduced. This is useful because the introduction of a true model improvement often produces a worse fit to data, since errors in other parts of the model are no longer compensated.

A product of the ensemble approach to parameter

estimation is a covarying distribution of parameter values. In this work we focused only on distributions of individual parameter values, but there is likely useful information in the higher moments of the parametric distribution. Additionally, we never expect our parameterization to be perfect, and the distribution of parameter values provides information about how parameters need to be altered to mimic inadequacies in the model. These distributions could potentially be exploited to develop stochastic parameterizations for use in ensemble forecast integrations.

In general, ensemble parameter estimation techniques are applicable to a wide range of other parameterizations, such as land surface (Jackson et al. 2003) or deep convective parameterizations (Emanuel and Žiković-Rothman 1999). Also, general circulation models may benefit in the future, when computational power has increased. A single cloud parameterization containing subparameterizations for each term is analogous to a single climate model containing different parameterizations for radiative transfer, gravity wave drag, and so forth. Both suffer the problem of compensating errors and model structural errors, and both may benefit from tools to diagnose those errors.

Acknowledgments. COAMPS is a registered trademark of the Naval Research Laboratory. J.-C. Golaz was supported by the Visiting Scientist Program at the NOAA/Geophysical Fluid Dynamics Laboratory, administered by the University Corporation for Atmospheric Research (UCAR). V. E. Larson, D. P. Schanen, and B. M. Griffin are grateful for financial support provided by Grant ATM-0442605 from the National Science Foundation, and by Subaward G-7424-1 from the DoD Center for Geosciences/Atmospheric Research at Colorado State University via Cooperative Agreement DAAD19-02-2-0005 with the Army Research Laboratory. J. A. Hansen acknowledges support from ONR YIP N00014-02-1-0473. C. Jackson is acknowledged for a useful discussion about this work.

APPENDIX

Model Predictive Equations

The SCM predictive variables consist of the grid-average horizontal winds ($\overline{u}$, $\overline{v}$), liquid water potential temperature ($\overline{\theta}_l$), and total water mixing ratio ($\overline{q}_t$). If the SCM was implemented as a parameterization, these four variables would already be predicted by the host model. In addition, the SCM carries all six second moments arising from w, θ_l, and q_t. A single third moment ($\overline{w'^3}$) is predicted. This type of model is sometimes referred to as incomplete third-order closure model. The governing equations are

$$\frac{\partial \overline{u}}{\partial t} = -\overline{w}\frac{\partial \overline{u}}{\partial z} - f(\upsilon_g - \overline{v}) - \frac{\partial}{\partial z}\overline{u'w'}, \tag{A1}$$

$$\frac{\partial \overline{v}}{\partial t} = -\overline{w}\frac{\partial \overline{v}}{\partial z} + f(u_g - \overline{u}) - \frac{\partial}{\partial z}\overline{v'w'}, \tag{A2}$$

$$\frac{\partial \overline{q}_t}{\partial t} = -\overline{w}\frac{\partial \overline{q}_t}{\partial z} - \frac{\partial}{\partial z}\overline{w'q_t'} + \left.\frac{\partial \overline{q}_t}{\partial t}\right|_{\text{ls}}, \tag{A3}$$

$$\frac{\partial \overline{\theta}_l}{\partial t} = -\overline{w}\frac{\partial \overline{\theta}_l}{\partial z} - \frac{\partial}{\partial z}\overline{w'\theta_l'} + \overline{R} + \left.\frac{\partial \overline{\theta}_l}{\partial t}\right|_{\text{ls}}, \tag{A4}$$

$$\begin{aligned}\frac{\partial \overline{w'^2}}{\partial t} = {} & -\overline{w}\frac{\partial \overline{w'^2}}{\partial z} - \frac{\partial \overline{w'^3}}{\partial z} - 2\overline{w'^2}\frac{\partial \overline{w}}{\partial z} + \frac{2g}{\theta_0}\overline{w'\theta_v'} \\ & - C_5\left(-2\overline{w'^2}\frac{\partial \overline{w}}{\partial z} + \frac{2g}{\theta_0}\overline{w'\theta_v'}\right) \\ & + \frac{2}{3}C_5\left(\frac{g}{\theta_0}\overline{w'\theta_v'} - \overline{u'w'}\frac{\partial \overline{u}}{\partial z} - \overline{v'w'}\frac{\partial \overline{v}}{\partial z}\right) \\ & - \frac{C_1}{\tau}\overline{w'^2} + \nu_1\nabla_z^2\overline{w'^2},\end{aligned} \tag{A5}$$

$$\begin{aligned}\frac{\partial \overline{q_t'^2}}{\partial t} = {} & -\overline{w}\frac{\partial \overline{q_t'^2}}{\partial z} - \frac{\partial \overline{w'q_t'^2}}{\partial z} - 2\overline{w'q_t'}\frac{\partial \overline{q}_t}{\partial z} - \frac{C_2}{\tau}\overline{q_t'^2} \\ & + \nu_2\nabla_z^2\overline{q_t'^2},\end{aligned} \tag{A6}$$

$$\begin{aligned}\frac{\partial \overline{\theta_l'^2}}{\partial t} = {} & -\overline{w}\frac{\partial \overline{\theta_l'^2}}{\partial z} - \frac{\partial \overline{w'\theta_l'^2}}{\partial z} - 2\overline{w'\theta_l'}\frac{\partial \overline{\theta}_l}{\partial z} - \frac{C_2}{\tau}\overline{\theta_l'^2} \\ & + \nu_2\nabla_z^2\overline{\theta_l'^2},\end{aligned} \tag{A7}$$

$$\begin{aligned}\frac{\partial \overline{q_t'\theta_l'}}{\partial t} = {} & -\overline{w}\frac{\partial \overline{q_t'\theta_l'}}{\partial z} - \frac{\partial \overline{w'q_t'\theta_l'}}{\partial z} - \overline{w'q_t'}\frac{\partial \overline{\theta}_l}{\partial z} - \overline{w'\theta_l'}\frac{\partial \overline{q}_t}{\partial z} \\ & - \frac{C_2}{\tau}\overline{q_t'\theta_l'} + \nu_2\nabla_z^2\overline{q_t'\theta_l'},\end{aligned} \tag{A8}$$

$$\begin{aligned}\frac{\partial \overline{w'q_t'}}{\partial t} = {} & -\overline{w}\frac{\partial \overline{w'q_t'}}{\partial z} - \frac{\partial \overline{w'^2q_t'}}{\partial z} - \overline{w'^2}\frac{\partial \overline{q}_t}{\partial z} - \overline{w'q_t'}\frac{\partial \overline{w}}{\partial z} \\ & + \frac{g}{\theta_0}\overline{q_t'\theta_v'} - \frac{C_6}{\tau}\overline{w'q_t'} \\ & - C_7\left(-\overline{w'q_t'}\frac{\partial \overline{w}}{\partial z} + \frac{g}{\theta_0}\overline{q_t'\theta_v'}\right) + \nu_6\nabla_z^2\overline{w'q_t'},\end{aligned} \tag{A9}$$

$$\frac{\partial \overline{w'\theta_l'}}{\partial t} = -\overline{w}\frac{\partial \overline{w'\theta_l'}}{\partial z} - \frac{\partial \overline{w'^2\theta_l'}}{\partial z} - \overline{w'^2}\frac{\partial \overline{\theta}_l}{\partial z} - \overline{w'\theta_l'}\frac{\partial \overline{w}}{\partial z} + \frac{g}{\theta_0}\overline{\theta_l'\theta_v'} - \frac{C_6}{\tau}\overline{w'\theta_l'} - C_7\left(-\overline{w'\theta_l'}\frac{\partial \overline{w}}{\partial z} + \frac{g}{\theta_0}\overline{\theta_l'\theta_v'}\right) + \nu_6\nabla_z^2\overline{w'\theta_l'}, \tag{A10}$$

$$\frac{\partial \overline{w'^3}}{\partial t} = -\overline{w}\frac{\partial \overline{w'^3}}{\partial z} - \frac{\partial \overline{w'^4}}{\partial z} + 3\overline{w'^2}\frac{\partial \overline{w'^2}}{\partial z} - 3\overline{w'^3}\frac{\partial \overline{w}}{\partial z} + \frac{3g}{\theta_0}\overline{w'^2\theta_v'} - \frac{C_8}{\tau}(C_{8b}\mathrm{Sk}_w^4 + 1)\overline{w'^3} - C_{11}\left(-3\overline{w'^3}\frac{\partial \overline{w}}{\partial z} + \frac{3g}{\theta_0}\overline{w'^2\theta_v'}\right) + (K_w + \nu_8)\nabla_z^2\overline{w'^3}, \tag{A11}$$

where $\overline{R}$ is the radiative heating rate, f is the Coriolis parameter, and u_g and v_g are the geostrophic winds. Here $(\partial \overline{q}_t/\partial t)|_{\mathrm{ls}}$ and $(\partial \overline{\theta}_l/\partial t)|_{\mathrm{ls}}$ are large-scale moisture and temperature forcings, respectively. Here g is the gravity, and ρ_0 and θ_0 are the reference density and potential temperature, respectively. For additional details, the reader is referred to Golaz et al. (2002a). The PDF functional form used to close the buoyancy and higher-order moments follows Larson and Golaz (2005). One exception is the treatment of the PDF parameter $\tilde{\sigma}_w^2$. Instead of using Eq. (37) in Larson and Golaz (2005), we keep $\tilde{\sigma}_w^2$ constant but include it in the list of parameters to be calibrated.

Some aspects of the numerical integration have changed compared to Golaz et al. (2002a). Many higher-order turbulence moments are now discretized semi-implicitly to improve numerical stability and allow for longer time steps. Equations (A6)–(A8) are now solved for their steady-state solutions, thus making them diagnostic. The form of the damping term in (A11) has been altered. The change does not impact the results significantly, but it allows for a semi-implicit numerical treatment, which improves stability.

REFERENCES

Aksoy, A., F. Zhang, and J. W. Nielsen-Gammon, 2006a: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Mon. Wea. Rev.,* **134,** 2951–2970.

——, ——, and ——, 2006b: Ensemble-based simultaneous state and parameter estimation with MM5. *Geophys. Res. Lett.,* **33,** L12801, doi:10.1029/2006GL026186.

Annan, J. D., J. C. Hargreaves, N. R. Edwards, and R. Marsh, 2005a: Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Modell.,* **8,** 135–154.

——, D. J. Lunt, J. C. Hargreaves, and P. J. Valdes, 2005b: Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter. *Nonlinear Processes Geophys.,* **12,** 363–371.

Brown, A. R., and Coauthors, 2002: Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quart. J. Roy. Meteor. Soc.,* **128,** 1075–1093.

Carrió, G., W. R. Cotton, and D. Zupanski, 2006: Data assimilation into a LES model: Retrieval of IFN and CCN concentrations. Preprints, *12th Conf. on Cloud Physics,* Madison, WI, Amer. Meteor. Soc., 1.4A.

Deardorff, J. W., 1980: Cloud top entrainment instability. *J. Atmos. Sci.,* **37,** 131–147.

Emanuel, K. A., and M. Živković-Rothman, 1999: Development and evaluation of a convection scheme for use in climate models. *J. Atmos. Sci.,* **56,** 1766–1782.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99,** 10 143–10 162.

Geman, S., E. Bienenstock, and R. Doursat, 1992: Neural networks and the bias/variance dilemma. *Neural Comput.,* **4,** 1–58.

Golaz, J.-C., V. E. Larson, and W. R. Cotton, 2002a: A PDF-based model for boundary layer clouds. Part I: Method and model description. *J. Atmos. Sci.,* **59,** 3540–3551.

——, ——, and ——, 2002b: A PDF-based model for boundary layer clouds. Part II: Model results. *J. Atmos. Sci.,* **59,** 3552–3571.

——, S. Wang, J. D. Doyle, and J. M. Schmidt, 2005: COAMPS®-LES: Model evaluation and analysis of second and third moment vertical velocity budgets. *Bound.-Layer Meteor.,* **116,** 487–517.

Hacker, J. P., and C. Snyder, 2005: Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized PBL. *Mon. Wea. Rev.,* **133,** 3260–3275.

Jackson, C., Y. Xia, K. Sen, and P. L. Stoffa, 2003: Optimal parameter and uncertainty estimation of a land surface model: A case study using data from Cabauw, Netherlands. *J. Geophys. Res.,* **108,** 4583, doi:10.1029/2002JD002991.

——, M. K. Sen, and P. L. Stoffa, 2004: An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions. *J. Climate,* **17,** 2828–2841.

Larson, V. E., and J.-C. Golaz, 2005: Using probability density functions to derive consistent closure relationships among higher-order moments. *Mon. Wea. Rev.,* **133,** 1023–1042.

——, ——, and W. R. Cotton, 2002: Small-scale and mesoscale variability in cloudy boundary layers: Joint probability density functions. *J. Atmos. Sci.,* **59,** 3519–3539.

Moeng, C.-H., and Coauthors, 1996: Simulation of a stratocumulus-topped planetary boundary layer: Intercomparison among different numerical codes. *Bull. Amer. Meteor. Soc.,* **77,** 261–278.

Moody, J., 1994: Prediction risk and architecture selection for neural networks. *From Statistics to Neural Networks: Theory and Pattern Recognition Applications,* V. Cherkassky, J. H. Friedman, and H. Wechsler, Eds., NATO ASI Series F, Vol. 136, Springer-Verlag, 143–156.

Pope, S. B., 2000: *Turbulent Flows.* Cambridge University Press, 771 pp.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flan-

nery, 1992: *Numerical Recipes in FORTRAN: The Art of Scientific Computing.* 2d ed. Cambridge University Press, 965 pp.

Randall, D. A., 1980: Conditional instability of the first kind upside-down. *J. Atmos. Sci.,* **37,** 125–130.

Siebesma, A. P., and Coauthors, 2003: A large eddy simulation intercomparison study of shallow cumulus convection. *J. Atmos. Sci.,* **60,** 1201–1219.

Stevens, B., and Coauthors, 2001: Simulations of trade wind cumuli under a strong inversion. *J. Atmos. Sci.,* **58,** 1870–1891.

——, and Coauthors, 2003: Dynamics and chemistry of marine stratocumulus—DYCOMS-II. *Bull. Amer. Meteor. Soc.,* **84,** 579–593.

——, and Coauthors, 2005: Evaluation of large-eddy simulations via observations of nocturnal marine stratocumulus. *Mon. Wea. Rev.,* **133,** 1443–1462.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.

Zhu, P., and Coauthors, 2005: Intercomparison and interpretation of single-column model simulations of a nocturnal stratocumulus-topped marine boundary layer. *Mon. Wea. Rev.,* **133,** 2741–2758.

Zupanski, D., and M. Zupanski, 2006: Model error estimation employing an ensemble data assimilation approach. *Mon. Wea. Rev.,* **134,** 1337–1354.